The Spatial Aspect of Color and Scientific Implications of Retinex Image Processing

Daniel J. Jobson[†], Zia-ur Rahman[‡], Glenn A. Woodell[†]

[†]NASA Langley Research Center, Hampton, Virginia 23681.
[‡]College of William & Mary, Department of Computer Science, Williamsburg, Virginia 23187.

ABSTRACT

The history of the spatial aspect of color perception is reviewed in order to lay a foundation for the discussion of retinex image processing. While retinex computations were originally conceived as a model for color constancy in human vision, the impact on local contrast and lightness is even more pronounced than the compensation for changes in the spectral distribution of scene illuminants. In the multiscale retinex with color restoration (MSRCR), the goal of the computation is fidelity to the direct observation of scenes. The primary visual shortcoming of the recorded image is that dark zones such as shadow zones are perceived with much lower contrast and lightness than for the direct viewing of scenes. Extensive development and testing of the MSRCR led us to form several hypotheses about imaging which appear to be basic and general in nature. These are: (1) the linear representation of the image is not usually a good visual representation, (2) retinex image enhancements tend to approach a statistical ideal which suggests the existence of a canonical "visual image", and (3) the mathematical form of the MSRCR suggests a deterministic definition of visual information which is the log of the spectral and spatial context ratios for any given image. These ideas imply that the imaging process should be thought of, not as a replication process whose goal is minimal distortion, but rather as a profound non-linear transformation process whose goal is a statistical ideal visual representation. These insights suggest new directions for practical advances in bringing higher levels of visual intelligence to the world of computing.

1. INTRODUCTION—THE SPATIAL ASPECT OF COLOR

Visual perception is replete with surprises. Among the most basic is that color perception has a strong spatial dependency rather than being purely spectral in nature. This was demonstrated dramatically by Edwin Land with color constancy experiments. By manipulating the color of light sources so that green and red reflectance patches were spectrally identical, he showed that they are still seen as green and red. In fact one would be hard pressed to define any explanation of human vision's color constancy which does not demand an interaction between spectral and spatial processing of images.

Land postulated a number of variations[1–4] of his retinex theory which culminated in a last version[3] taking the form of a non-linear center/surround. This form was the starting point for our MSRCR which elaborates the concept to include multiple scales of surrounds and defines a color restoration which overcomes the fundamental practical limitation of the gray-world assumption intrinsic to the retinex concept. Multiple scales and the color restoration were required to produce an general purpose and automatic image enhancement method with graceful tonal and color performance.

More recently, a striking new observation has arisen from a study of the color perception of the optical spectra. Smeulders et al.[5] noted discrepancies in the historical data of various key figures in color physics—Newton, Young, and Helmholtz. These all centered of differing descriptions of the colors seen in spectra projected from prisms and were traced to differences in the manner of demarcation used. A thorough

DJJ: d.j.jobson@larc.nasa.gov; co-authors: ZR: zrahman@cs.wm.edu; GAW: g.a.woodell@larc.nasa.gov

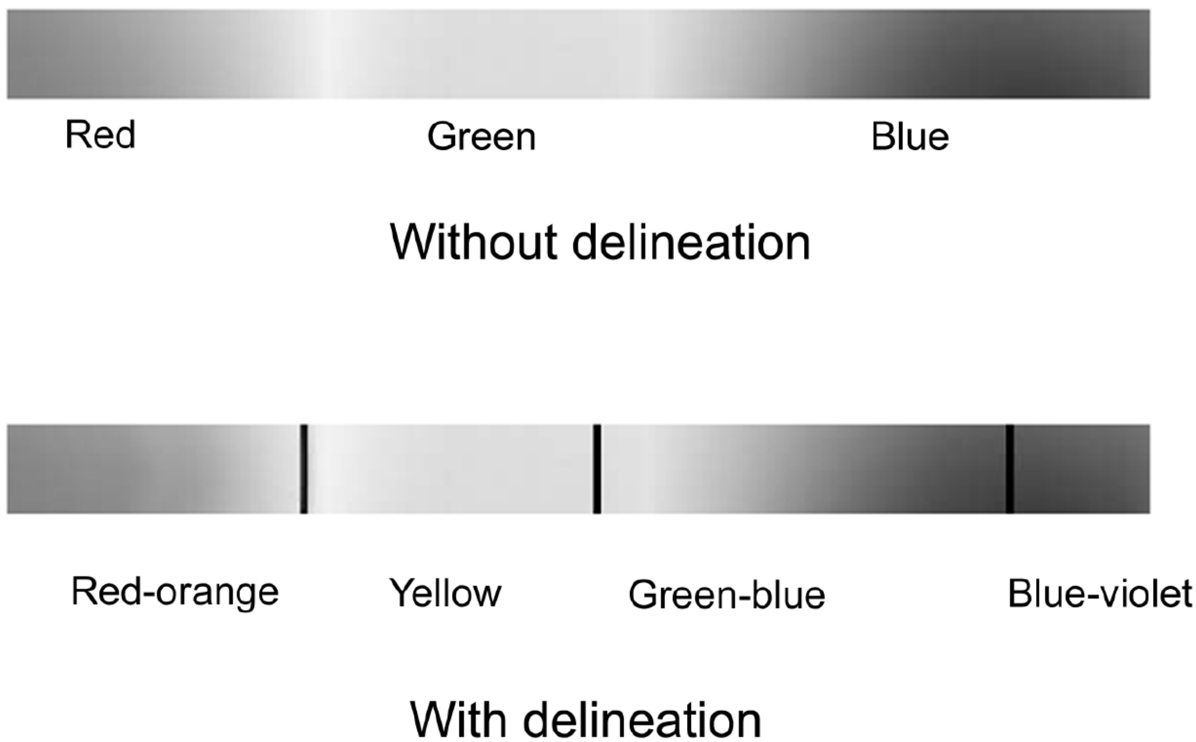

Figure 1. Perception of spectral color

study of the perception of spectral color led to the conclusion that only three primary colors are seen in a spectrum without delineation (Figure 1). When lines are added to the spectrum, the full color gamut begins to emerge- more lines, more colors up to a saturation point where the addition of more lines brings about fewer and fewer additional hues. The implication like that of Land's work is that there is a strong influence of spatial processing in color perception. In this case, the conclusion is very basic—that spatial structure is essential for full color perception.

2. RETINEX IMAGE PROCESSING—CONTRAST AND LIGHTNESS FIDELITY TO DIRECT VIEWING OF SCENES

Our own work with retinex image processing[6,7] leads us away from the world of color and into the world of contrast/lightness perception of complex natural scenes. While the MSRCR synthesizes color constancy, dynamic range compression, and the enhancement of contrast and lightness—the emphasis here is on the latter. We have used the MSRCR with many tens of thousands of test images and find it to be a generic image enhancement computation. Specifically, it does bring the perception of dark zones in recorded images up in local lightness and contrast to the degree needed to mimic direct scene viewing. Only images with very modest dynamic ranges do not need enhancement and for these the exposure must be very accurate to achieve a good visual representation. Even wide ranging reflectance values in a scene, and certainly strong lighting variations demand a rather strong enhancement to achieve any thing like the visual realism of direct observation. The dynamic range compression of the retinex computation is the basis for the contrast/lightness enhancement. The generic character of the computation is the basis for using it as an automatic enhancement. A few examples of retinex enhancements will serve to convey the degree to which images need to be improves and provide a demonstration that the MSRCR does, in fact, perform this task

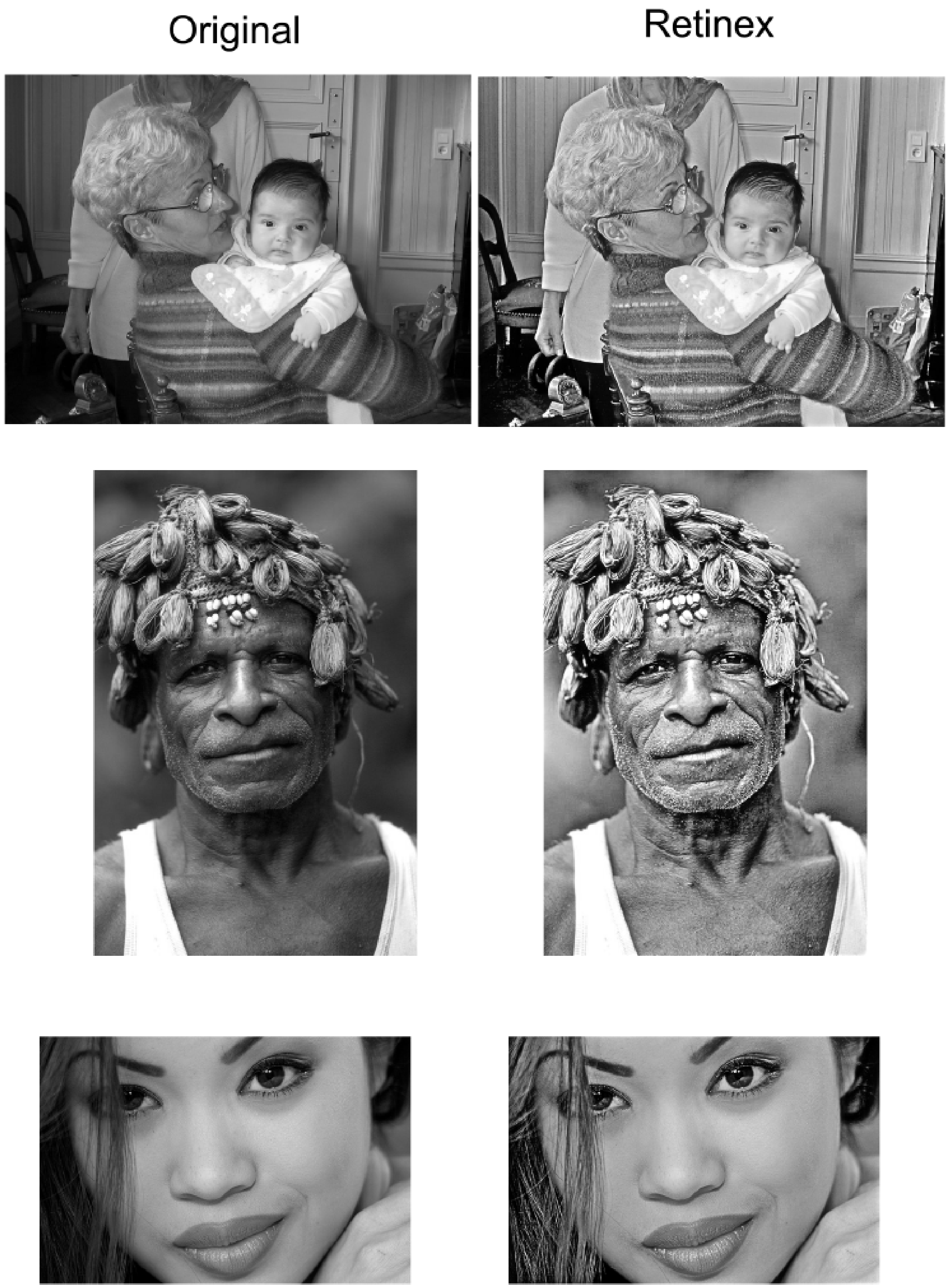

Figure 2. Retinex examples to illustrate that the strength of the enhancement matches the degree of visual deficit in the original image. (a) Subtle enhancements

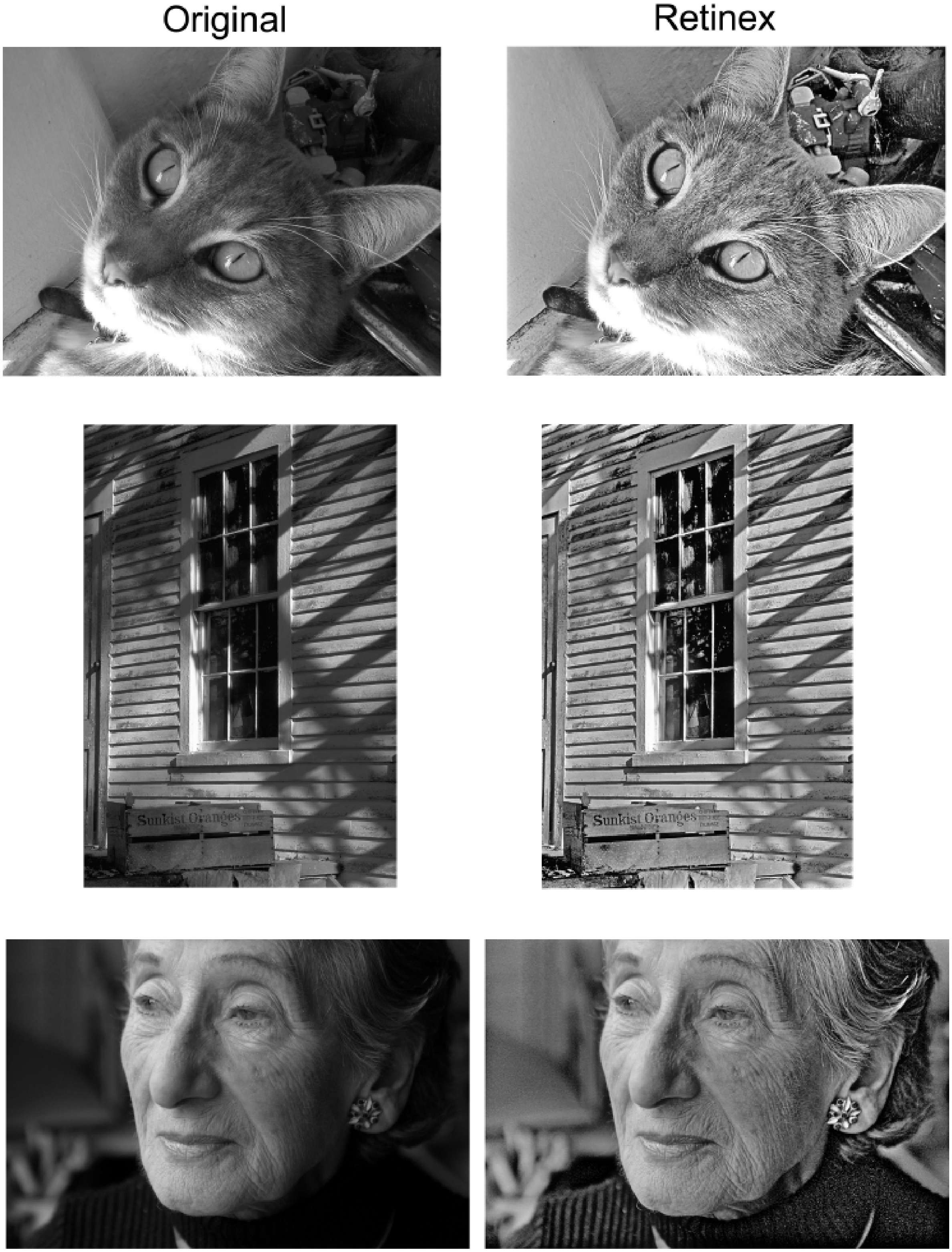

Figure 2. Continued: (b) Moderate enhancements

Figure 2. Continued: (b) Stong enhancements

with considerable agility (Figure 2) and without human intervention. These examples highlight a major facet of retinex performance: intrinsically, the degree of automatic enhancement matches the degree of visual deficit in the original acquired image.

The design of the MSRCR was defined by experiments whereby computational parameters were adjusted until the processed image display or print compared well with the direct viewing of test scenes. This was then tested successfully on a large very diverse battery of several thousand test images in order to ferret out any quirks in performance or rare pathologies. Since most digital images do not come with any knowledge of pre-processing which may have been applied at the camera or scanner software driver level or by subsequent image processing tools, the impact of retinex performance on various commonplace forms of preprocessing was studied. While distortions were observed, they were generally mild and the MSRCR was reasonably resilient to a range of gamma and contrast stretch operations applied prior to the retinex computation.[8] Ideally the retinex should be applied to linear data with only dark offsets removed and with no image compression or high quality levels of JPEG. The impact of preprocessing (either gamma or contrast stretch) is to shift the optimum post-retinex gain/offset values. Should preprocessing be consistent, then a resetting of the post-retinex gain/offset can be used to produce best performance on preprocessed image data.

While the MSRCR was developed with general purpose color imaging in mind, it solves such a fundamental problem of imaging—good visibility across a wide dynamic range of data—that it is useful for scientific and other special purpose imaging applications such as medical, forensic, surveillance, and remote sensing applications (Figure 3). It has been shown in previous studies[9,10] to improve multispectral classification accuracies when used as as a front-end to supervised training classification schemes where extensive ground truth was available. The practical value of retinex image processing is accompanied by the emergence of new scientific insights into the visual image and the imaging process. These will be outlined in the remainder of this text.

3. SCIENTIFIC IMPLICATIONS OF THE RETINEX EXPERIENCE

3.1. The visual inadequacy of the linear representation

During the course of developing this retinex computation and testing it experimentally on large numbers of diverse images, we were forced to re-examine some of our most basic ideas about the imaging process and found that some were no longer tenable. If we assume that the goal of imaging is a good visual representation which compares to the direct observation of scenes for color imaging or simply provides good visibility for imaging outside visual spectral range (IR, MRI, Xray,etc.), we had to discard the idea that imaging is a replication process whose goal is minimal distortion of measured signals or radiometry. In place of this long-standing tradition of much of signal and image processing, we had to move to the idea that imaging is a process of profound transformation which intrinsically involves non-linear spatial processing. This shift arises entirely from considering the image as a visual entity and the evident visual shortcomings of the linear representation of image data (Figure 4). In general the linear representation is not a good visual representation. This is consistent with the conclusion of a study of the data handling and processing for color negative film scanning.[11] Tuijn describes the correction for all transfer functions so that the image data is linear, and then explains that this is often visually inadequate—weak in contrast and color. In order to explore this further, we displayed known linear data taken with a Nikon D1 camera in linear mode on linearized color computer monitor (gamma correction of 1.6). For a wide array of images, the displayed image is too dark (Fig 4), and the retinex enhancement (also shown for comparison) was required to produce a good visual representation. The linear representation can approach a good visual rendering for a very restricted class of scenes—those with diffuse illumination and restricted ranges of

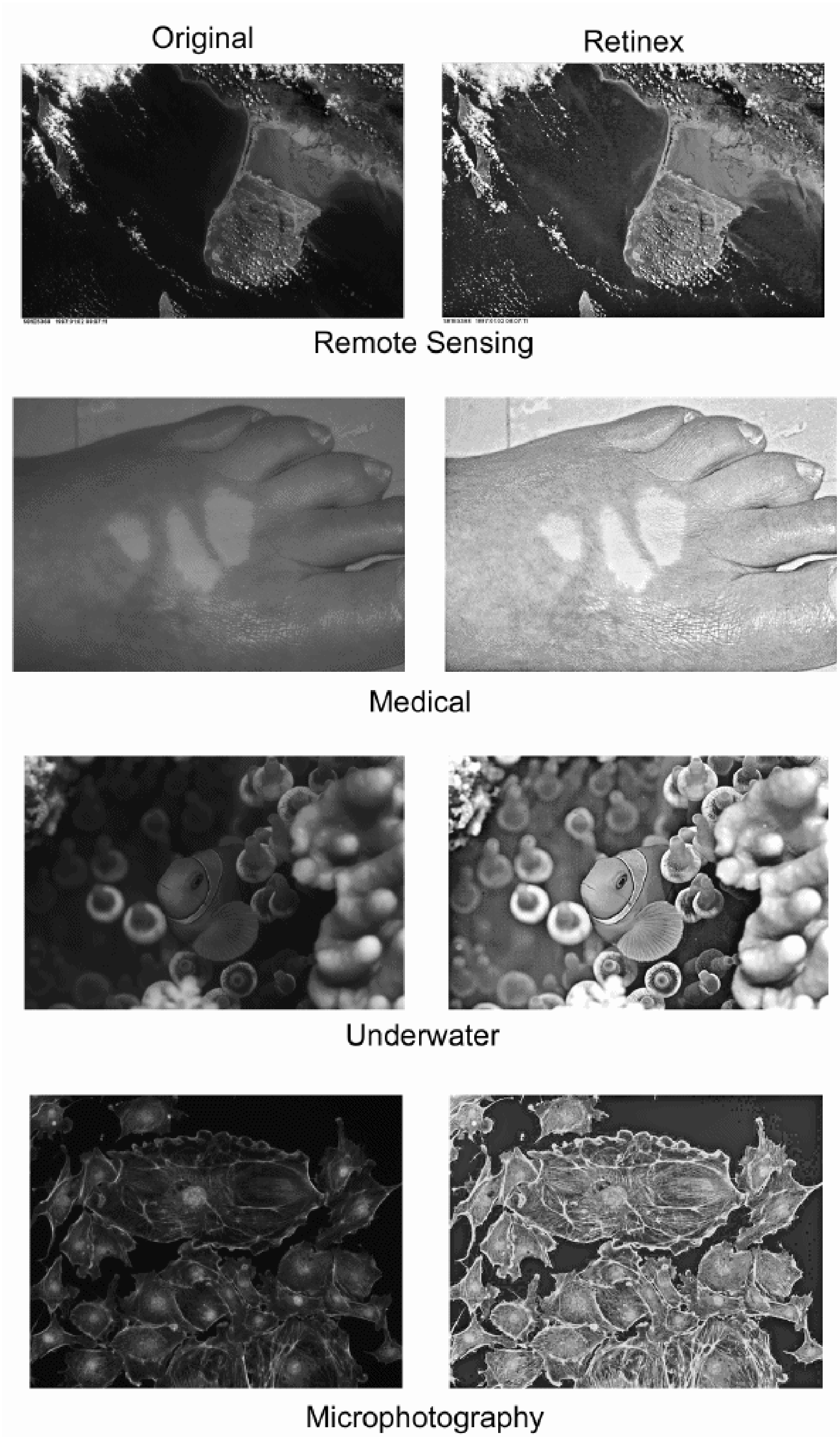

Figure 3. Illustration of specialized applications

Figure 4. Visual inadequacy of the linear representation

reflectances (or those where white surfaces which can be saturated). Even so, for this cooperative class a substantial degree of contrast stretching (gain/offset) is required to achieve a good visual display/print.

3.2. The existence of a canonical visual image and definition of visual measures for automating visual assessment

While image data is quite arbitrary in a statistical sense, we observe that retinexed data were not. As noted in a previous paper,[7] retinex histograms tend toward a characteristic Gaussian-like shape. More recently we have studied regional means (visual lightness) and standard deviations (visual contrast) and found that they tend to converge on consistent global aggregates. This implies that a good visual representation can be associated with well-defined statistical measures for visual quality. In scientific terms, this implies the existence of a canonical visual image as a statistical practical ideal. Such a defined ideal can then serve as the basis for the automatic assessment of visual quality. While this work is still underway, we can show some preliminary results which are encouraging. By following the general idea that the retinex brings regional means and standard deviations up to higher values and that these approach an ideal goal, we have constructed tentative visual measures and performed some testing. The measures were set empirically on a small diverse test image set and then were applied to a broad array of images of all sorts. Figure 5(a) shows a sample of the automatic visual quality assessment by classification into one of three classes—poor, good, excellent. The classification scheme is based upon the map shown in Figure 5(b).

While more study and development is necessary, the early results do support the idea of a canonical visual image with well defined statistical properties. Further, the investigation indicates that the MSRCR is a valuable tool for research purposes—in this case, to define a new statistical measure of visual quality.

Currently, computers have essentially no visual intelligence. Visual measures which can enable the automatic assessment of digital images are a first step in visual intelligence. Such measures would allow the computer to determine visual quality and automate image processing at a higher level and in a more sophisticated manner than is now possible.

3.3. A Hypothetical Deterministic Definition of Visual Information

While the retinex experience provides new avenues for the study for statistical image processing, it also suggests deterministic pathways as well. The generic character of the retinex computation suggests that some new quantitative definition of visual information may be possible. A deterministic definition would contrast with previous statistical ones based upon information theory.[12,13] Specifically the retinex is approximately performing a log of the ratio of each pixel in each spectral band to both spatial and spectral averages. The suppression of spatial and spectral lighting variations is achieved at the expense of accepting a significant degree of context dependency. Simply put the retinex mimics human perception in producing color and lightness which which are influenced by the visual setting in which they occur. The exchange of spatial and spectral lighting dependencies for spatio-spectral context effects appears to be a very basic element of human vision and the retinex computation. While we do not have a clear definition of information in a semantic sense, or visual information as some subset of all information, the idea that information is context relationships is appealing. The additional factor of a log function suggests a compactness which may be leading in the direction of symbolic representation—the symbol being the ultimate conciseness and carrier of meaning.

The establishment of context relationships is central to at least the senses of vision and hearing. Music seems to be based upon pitch relationships with certain ratios producing consonance or dissonance in varying degrees. Speech recognition must contend with the difficulties of speaker variations, the interdependencies of phonemes, and all manner of extraneous variations in loudness, temporal rates, degrees of clarity, and the like. For vision, the awesome task of transforming the signals of vision into the sense of

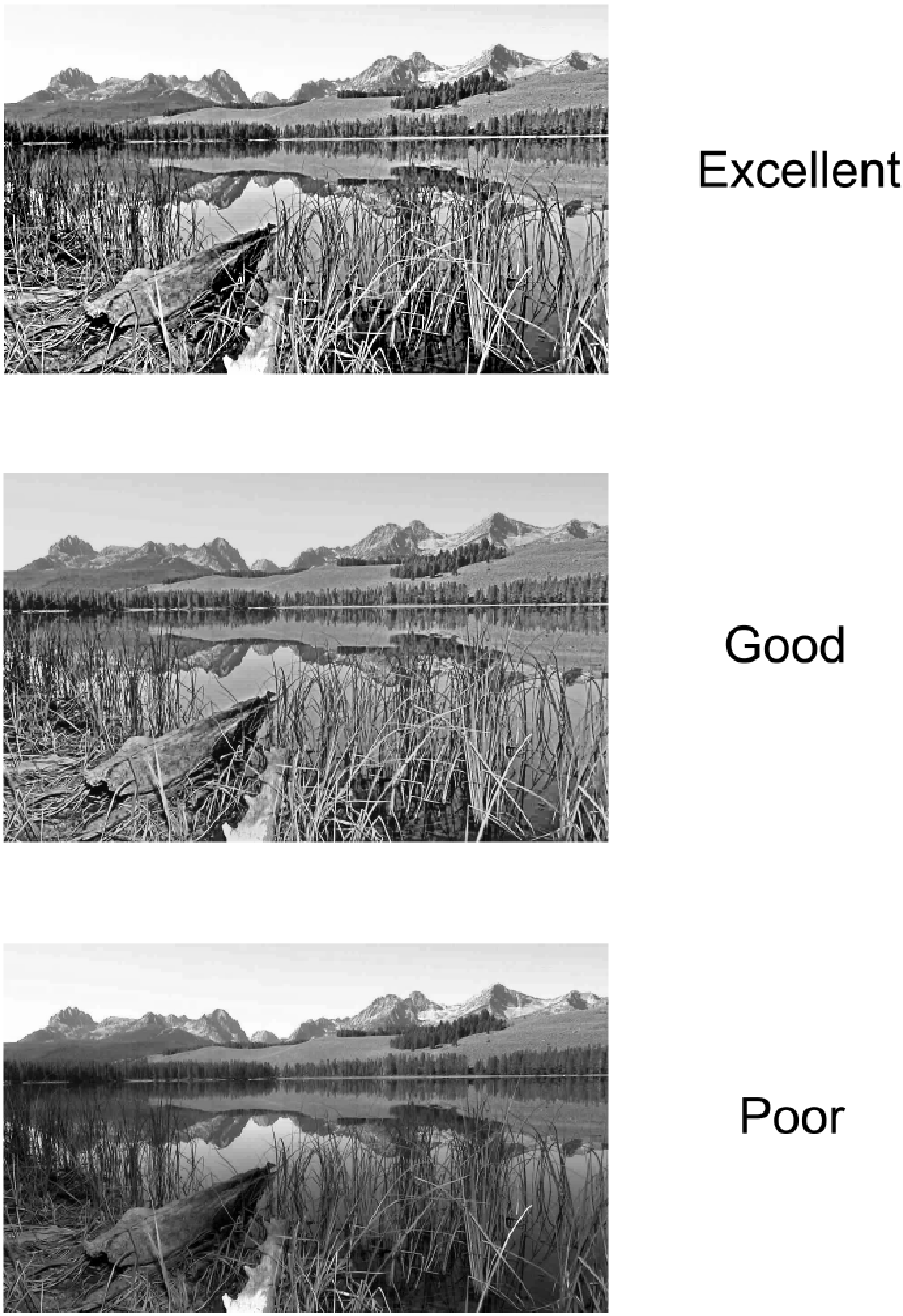

Figure 5. Preliminary performance of Visual Measures for automating visual assessment (a) Global classes

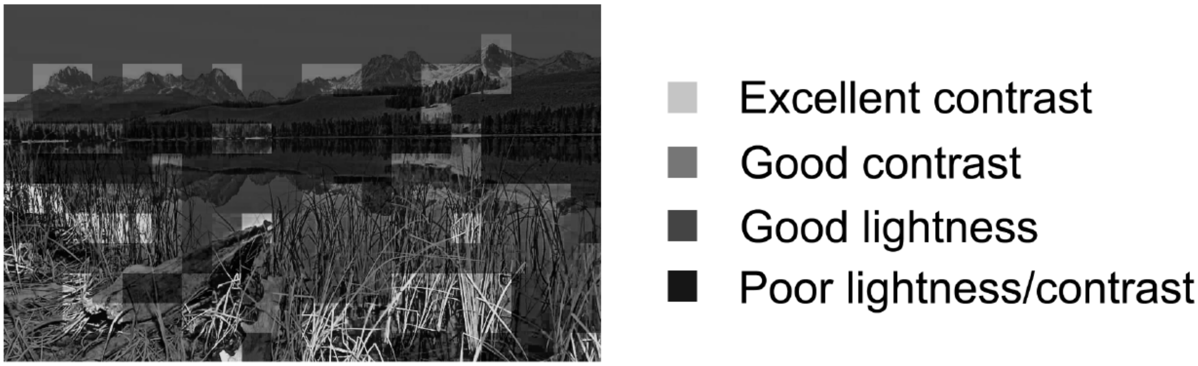

Figure 5. Continued: (b) Visual map showing regional classes

vision must succeed in extracting information in the presence of all manner of extraneous variations as well as find some very concise ultimately symbolic representation. Context must be a critical element of vision information as it is in speech and music where isolated acoustical events become perceived as a fluid temporal mesh of meaningful words or melody, harmony, and rhythm. Signals are not meaningful in isolation and for vision such contextual relationships as edge connectedness, textural uniformity, and color reflectances differences seem fundamental to building a some sort of "visual information". Perhaps the retinex transformation moves one step in this direction by reducing extraneous variations, increasing spatial and spectral differences, and providing a foundation for a structure of relatedness which with subsequent processing can become symbolic.

3.4. Impact of Retinex Processing on Sensor Design and End-to-End Processing

How should the retinex computation fit within the system of acquiring, processing, and displaying images? How does it exert an influence, if any, on sensor design and other image processing? With respect to sensor design the retinex from our practical experience, as well as from the consideration of scene radiometry, impacts the sensor design by asking for wide dynamic range image data with high signal-to-noise ratios (SNR). In order to minimize the visual distraction of emphatic noise in the retinex enhancement, wide dynamic range scenes should be acquired with high sensitivity sensors. For visible color imaging, image data dynamic ranges (at equivalent SNR) should (ideally) be in the 10–12bits range in order to encompass the scene radiometry of many everyday scenes and avoid noise visibility in the retinex enhancement of images with large very dark regions.

Ideally the retinex computation should be performed immediately after image acquisition and prior to other processing, especially data compression. Less ideally, the retinex can be applied with good results after compression if high quality levels (low compression ratios) of JPEG, for example, are used. There is, however a basic tension between the retinex computation and JPEG, in that JPEG hides artifacts in imperceivable dark zones which the retinex enhances into visibility when higher compression ratios are selected.

If the retinex is applied prior to compression by being embedded in a sensor chip or is applied to uncompressed data, the enhancement makes full use of the sensor performance envelope to achieve results that are limited only by the sensor performance itself—primarily SNR.

A front-end retinex has the advantage of bringing out the visual quality so that subsequent compression is done on this improved starting point. This reduces the likelihood that significant visual features will be distorted by JPEG or other coding artifacts.

4. CONCLUSION

The visual image remains an enigma full of surprises, some of which we encountered in our experience with retinex image processing. Our thinking about the imaging process has been changed in basic ways outlined here. The new directions stimulated are summarized as:

1. Imaging should be considered as a process of transformation rather than replication with minimal distortion. This idea arose from the visual inadequacy of the linear representation of image data.
2. The statistical convergence of retinex image enhancements led us to postulate the existence of a canonical visual image with consistent statistical aggregate characteristics. Further these can be used to construct entirely new visual measures which can be the basis for the automatic assessment of visual quality of arbitrary images by the computer.
3. A new deterministic definition of visual information emerges from the computational form of the retinex—namely that visual information is in some sense the log of spatial and spectral context relationships within the image.
4. The retinex should be applied ideally as a front-end computation prior to any data compression and achieves its fullest visual quality when sensor SNR and digitization cover the 10–12bits dynamic range required by the radiometric character of commonplace scenes.

A computation like the retinex appears to solve two problems simultaneously—the diminishing of extraneous variables such as spatial and spectral lighting and the construction of compact context relationships which may provide a basis for bringing more advanced levels of visual intelligence into computing.

REFERENCES

1. E. Land, "Color vision and the natural image," *Proc. Nat. Acad. Sci.* **45**, pp. 115–129, 1959.
2. E. Land, "Recent advances in retinex theory and some implications for cortical computations," *Proc. Nat. Acad. Sci.* **80**, pp. 5163–5169, 1983.
3. E. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," *Proc. Nat. Acad. Sci.* **83**, pp. 3078–3080, 1986.
4. E. Land, "Recent advances in retinex theory," *Vision Research* **26**(1), pp. 7–21, 1986.
5. N. Smeulders, F. W. Campbell, and P. R. Andrews, "The role of delienation and spatial frequency in the perceptoin of the colours of the spectrum," *Vision Research* **34**(7), pp. 927–936, 1994.
6. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. on Image Processing: Special Issue on Color Processing* **6**, pp. 451–462, March 1997.
7. D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multi-scale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing: Special Issue on Color Processing* **6**, pp. 965–976, July 1997.
8. Z. Rahman, D. Jobson, and G. A. Woodell, "Resliency of the multiscale retinex image enhancement algorithm," in *Proceedings of the IS&T Sixth color Imaging Conference: Color Science, Systems, and Applications*, pp. 129–134, IS&T, 1998.
9. B. D. Thompson, Z. Rahman, and S. K. Park, "A multi-scale retinex for improved performance in multi-spectral image classification," in *Visual Information Processing IX*, Proc. SPIE 4041, 2000.
10. B. D. Thompson, Z. Rahman, and S. K. Park, "Retinex preprocessing for improved multi-spectral image classification," in *Visual Information Processing VIII*, Proc. SPIE 3716, 1999.

11. C. Tuijn, "Scanning color negatives," in *Proceedings of the IS&T Fourth color Imaging Conference: Color Science, Systems, and Applications*, pp. 33–38, IS&T, 1996.
12. F. O. Huck, C. L. Fales, and Z. Rahman, "Information theory of visual communication," *Philosophical Transactions of the Royal Society of London A* **354**, pp. 2193–2248, Oct. 1996.
13. F. O. Huck, C. L. Fales, and Z. Rahman, *Visual Communication: An Information Theory Approach*, Kluwer Academic Publishers, Boston, MA, 1997.